

Durham Research Online

Deposited in DRO:

01 May 2019

Version of attached file:

Accepted Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Goff, P. (2016) 'Fundamentality and the mind-body problem.', *Erkenntnis*, 8 (4). pp. 881-898.

Further information on publisher's website:

<https://doi.org/10.1007/s10670-015-9773-7>

Publisher's copyright statement:

This is a post-peer-review, pre-copyedit version of an article published in *Erkenntnis*. The final authenticated version is available online at: <https://doi.org/10.1007/s10670-015-9773-7>

Additional information:

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

Fundamentality and the Mind-Body Problem

I: Introduction

In the recent metaphysics literature, a number of philosophers have independently endeavoured to marry sparse ontology to abundant truth. The aim is to keep ontological commitments minimal, whilst allowing true sentences to quantify over a vastly greater range of entities than those which they are ontologically committed to. For example, an ontological commitment only to concrete, microscopic simples might be conjoined with a commitment to truths such as ‘There are twenty people working in this building’ and ‘There are prime numbers greater than 5.’ Quinean orthodoxy tells us that if a sentence quantifies over people and numbers, then its truth can only be bought with an ontological commitment to people and numbers. But these radicals want more of a bargain: truths bought more cheaply than their quantificational price tags suggest is possible.¹

The aim of making sense of this kind of ‘cheap truth’ is to reconcile two powerful theoretical pulls. On the one hand Occam’s Razor pulling in the direction of parsimony, and on the other hand common sense pulling us in the direction of profligacy. The taste for desert landscapes is in tension with the Moorean resistance to abandoning our pre-theoretical commitments to tables, people, planets, numbers and sets. If we can make sense of cheap truth, then we can have our cake and eat it. Neither parsimony nor common sense need be abandoned.

Discussion and defence of cheap truth has taken place almost wholly within metaphysics, and has had little interaction with a more familiar and longer running issue in philosophy of mind, namely the debate between physicalists and anti-physicalists about consciousness. The aim of the paper is to press the importance of interaction between these two areas of the philosophical literature. I will present a significant challenge for the defenders of cheap truth theory, which has its roots in discussions of the mind-body problem.

Anti-physicalists have arguments to the conclusion that the truths about consciousness cannot be accounted for in physicalist terms. Those arguments themselves do not threaten cheap truth theory, as there are forms of anti-physicalism consistent with cheap truth theory. However, I will propose a new argument, using the key premises of the standard anti-physicalist arguments, to the conclusion

¹ Cameron 2008, 2010, Horgan and Potrč 2008. Sider 2009, 2012a, 2012b Williams 2012. Fine 2001 is an important influence on many of these views, but is not himself a cheap truth theorist.

that truths about conscious subjects cannot be accounted for in terms of the minimal ontological commitments favoured by the defenders of cheap truth. I will not defend the premises of that argument; there is enough discussion of this already in the philosophy of mind. But if this argument is valid, then cheap truth theorists are obliged to engage in the debates in the philosophy of mind concerning the truth-values of those key premises.

2: Horgan and Potrč's account of cheap truth

Let us define 'cheap truth theories' as those which are committed to the following two theses:

Miminal Commitments – All objects in the correct ontology are concrete and simple.

Moorean Truth – Scientifically informed common sense is largely correct, in the sense that our ordinary talk of composite objects, such as tables and planets, and abstract objects, such as sets and numbers is mostly true (or at least, to the degree that it is significantly false, this is the result of scientific, not metaphysical, ignorance).

A number of proposals for making sense of cheap truth theory have been made in recent years, and I think that the challenge I wish to present applies to all of them. However, I shall focus on the version of cheap truth theory articulated by Terry Horgan and Matjaz Potrč, and construe cheap truth theory in their terms.²

Horgan and Potrč take truth to be *semantically correct affirmability under contextually operative semantic standards*. For a sentence to be semantically correct it must correspond to the world. However, there are two ways in which a sentence might correspond to the world: directly or indirectly. For a sentence to directly correspond to the world is roughly for its structure to mirror the structure of the corresponding state of affairs, i.e. the entities it quantifies over correspond to entities in the world, its predicates correspond to properties/relations in the world, and the relevant entities stand in relations/instantiate properties in the way that is suggested by the sentence's superficial grammar.³ The sentence 'God is angry' directly corresponds to the world if the world contains God instantiating the property of being angry. A sentence indirectly corresponds to the world just in case it corresponds to the world in a way that is not direct. The sentence 'The table is near to the chair' indirectly corresponds to the world if it corresponds to the state of affairs of some particles arranged table-wise being located near to some particles arranged chair-wise.

² Horgan and Potrč 2008.

³ Although this goes beyond what Horgan and Potrč suggest, it may be possible to give an account of direct correspondence which does not require predicates to correspond to properties, and which would thereby be compatible with austere nominalism. I will not explore this here.

Horgan and Potrč suggest that most of the time the contextually operative standards which govern semantic correctness require only an indirect correspondence between sentence and world. If we are discussing how best to arrange the living room for a party, and I say, ‘Well, the table is near to the chair’, the semantic standards in operation in that conversation do not entail that the semantic correctness of the sentence requires a direct correspondence, i.e. that in the relevant bit of reality there is a table located near to a chair. An indirect correspondence to particles arranged table-wise located near particles arranged chair-wise will suffice to make the sentence semantically correct.

In relatively few contexts does semantic correctness require a direct correspondence between sentence and world, according to Horgan and Potrč. The most obvious case is when we are practicing ontology. In the metaphysics seminar room, the sentence ‘The table is near to the chair’ is semantically correct only if, in the relevant bit of reality, there exists a table and there exists a chair, and the table is located near to the chair. It may also be that expressions of our theological commitments require direct correspondence in order to be semantically correct: ‘God is perfectly good’ is arguably true only if God exists and has the property of being perfectly good.⁴

Separating out these two contexts allows Horgan and Potrč to embrace both ontological austerity and commonsense profligacy. Inside the ontology room, truth sentences quantify over simple, concrete objects. Outside the ontology room, true sentences quantify over the abundance of commonsense objects: tables, rocks, numbers, sets, etc. The semantic correctness and thereby truth, of sentences uttered outside the ontology room, is guaranteed by their indirect correspondence to facts about the simple concrete objects quantified over in the ontology room.

Cheap truth theories, as I have defined them, involve the thesis of *Minimal Commitments*. For most of this paper I am going to construe *Minimal Commitments* as *micro-simplism*, according to which we are ontologically committed only to microscopic simples (or ‘micro-simples’ for short). This is the most straightforward way of understanding *Minimal Commitments*, and it is the way most cheap truth theorists construe it.

⁴ It seems that Horgan and Potrč implicitly assume some metaphysical privileged notion of ‘Existence’ or ‘Reality’, such that the objects and properties to which sentences correspond Exist or are Real in this sense. For if they were just working with our everyday notion of ‘existence’, it would be trivial that ‘The table is near to the chair’ corresponds to a table, as the sentence trivially implies that a table exists (at least in our ordinary sense of ‘exists’). Most of the cheap truth theorists seem to reply on some such implicitly distinction between ordinary existence/reality and metaphysically significant Existence/Reality, such that it is the things we take to Exist that constitute our metaphysical commitments. In his 2010 paper Cameron makes this explicit by appealing to Sider’s (2008/2012) notion of quantificational structure, and I have confirmed with Williams in conversation that he intends his talk of what worlds ‘represent to be the case’ (Williams 2010) to be understood as what obtains at a world in a metaphysically serious sense. Fine (2001) and Sider (2009, 2012) are more explicit, giving detailed accounts of how we are to distinguish between the way we carve the world and the way the world is carved in and of itself.

However, although Horgan and Potrč give considerable credence to micro-simplism, their preferred view is the even more austere view that *only one object* is quantified over in the ontology room: the world itself, or as they call it the ‘bobject’. The bobject, by instantiating its various properties in spatio-temporally local manners, indirectly corresponds to our ordinary sentences. By instantiating various physical properties in table-ish and chair-ish spatio-temporal manners, the bobject can make true the sentence, ‘The table is near to the chair’. I will consider blobjectivist versions of *Minimal Commitments*, and hence of cheap truth theory, after I have considered micro-simplist versions.

3: The anti-physicalist arguments and phenomenal transparency

By ‘the anti-physicalist’ arguments I shall mean those arguments which try to establish an epistemic gap between the physical and the mental, and then to infer from that epistemic gap to a metaphysical gap inconsistent with the truth of physicalism. The zombie conceivability argument begins by trying to establish that zombies are conceivable. The first stage of the knowledge argument is to establish that Mary learns something new when she leaves her black and white room. The explanatory gap argument begins with some sense in which conscious states cannot be explained in terms of physical states.⁵ Each of these constitutes an interesting epistemic gap between the physical facts and the facts about consciousness.

Each of these arguments then tries to establish some thesis about the special nature of phenomenal concepts, i.e. our concepts of conscious states qua conscious states, which is involved in facilitating the move from the epistemic gap to a modal or metaphysical gap. David Chalmers argues that the primary intension of a phenomenal concept is the same as its secondary intension. George Bealer focuses on the *semantic stability* of phenomenal concepts: roughly the fact that they’re not ‘twin-Earthable.’⁶ Martine Nida-Rümelin has argued for property dualism on the basis of the fact that we *grasp* phenomenal qualities in deploying phenomenal concepts.⁷ And I have argued against physicalism on the basis that phenomenal concepts are *transparent*, i.e. that a phenomenal concept reveals the essence of the phenomenal quality it refers to.⁸

The challenge I wish to put to cheap truth theories could be articulated with the premises of any of the arguments mentioned above. However, for the sake of keeping things simple I will focus on Chalmers’ two-dimensional argument against materialism, which goes as follows (where *P* is the

⁵ Chalmers 1996, 2002, 2009, Jackson 1982, 1986, Levine 1983.

⁶ Bealer 1994, 2002.

⁷ Nida-Rumelin 2007

⁸ Goff 2011.

complete physical truth, and Q is an arbitrary truth about consciousness, such as 'Something is in pain'):

- (1) $P \& \sim Q$ is conceivable
- (2) If $P \& \sim Q$ is conceivable, then $P \& \sim Q$ is 1-possible
- (3) If $P \& \sim Q$ is 1-possible, then $P \& \sim Q$ is 2-possible or Russellian monism is true.
- (4) If $P \& \sim Q$ is 2-possible, physicalism is false.
- (5) Physicalism is false or Russellian monism is true.⁹

Notice that premise 3 opens up a loophole for the physicalist: the adoption of *Russellian monism*. Understanding this loophole is important for understanding why the anti-physicalist arguments in themselves are not a threat to cheap truth theory, and hence for understanding why the arguments need to be reformulated in the manner I suggest in this paper if they are to pose a challenge to cheap truth theory. However, I will save this discussion for the next section; for the moment we can ignore this loophole and focus on a formulation of the argument Chalmers considers earlier in the paper before raising the issue of Russellian monism:

- (1) $P \& \sim Q$ is conceivable
- (2) If $P \& \sim Q$ is conceivable, then $P \& \sim Q$ is 1-possible
- (3) If $P \& \sim Q$ is 1-possible, then $P \& \sim Q$ is 2-possible.
- (4) If $P \& \sim Q$ is 2-possible, physicalism is false.
- (5) Physicalism is false.

Understanding the argument requires a little understanding of Chalmers' two-dimensional semantic framework, which I shall try to convey before turning to the argument.

We can model the content of a concept as a function from possible world to referents (in what follows I will refer to concepts with underlined terms). For example, we can think of the concept president of the USA as a function that delivers at each world the thing that is the president of the USA in that world: delivers Obama in the actual world, delivers Romney in the possible world where Romney won the election, etc. Some concepts deliver the same referent at each possible world, or

⁹ Chalmers 2009.

at least at each world at which they deliver anything at all. For example, water delivers H₂O at each possible world at which it refers.

Similarly, we can model sentences as functions from possible worlds to truth-values. 'Obama is president of the USA' delivers the truth-value *true* at the actual world, but the truth-value *false* at the world where Romney won the election. 'Water is H₂O' delivers the truth-value *true* at every possible world. These functions of concepts and sentences are known as 'intensions'.

The essence of the two-dimensional framework is the supposition that each concept/sentence is associated with *two* functions from worlds to referents/truth-values.¹⁰ In the examples above, I was giving what Chalmers calls the 'secondary intension' of water, president of the USA and of sentences comprising terms which express those concepts. But Chalmers thinks these concepts and sentences are also associated with a *primary intension*.

The difference between the primary and secondary intension is given by the way in which worlds are considered when evaluating the intension. When we are evaluating the primary intension, we consider worlds *as actual*; when we are evaluating the secondary intension we consider worlds *as counterfactual*. I will try to explain both these modes of evaluation, and the resulting distinction between primary and secondary intensions, in terms of 'the XYZ world': the world in which the colourless, odourless stuff in oceans and lakes is XYZ and H₂O.

To consider a world as actual is to consider it as a way the actual world might turn out to be. When we consider the XYZ world as actual, we are entertaining the supposition that the actual colourless odourless stuff in oceans and lakes is XYZ rather than H₂O. On this supposition, water refers not to H₂O but to XYZ, and 'Water is H₂O' is not true but false. We can say, then, that the primary intension of water delivers the referent XYZ at the XYZ world, and the primary intension of water is H₂O delivers the truth-value false at the XYZ world.¹¹

To consider a world as counterfactual is to consider it as a way things might have been. When we consider the XYZ world as actual, we bear in mind that the actual world is one where the colourless, odourless stuff in oceans and lakes is H₂O, and we consider the XYZ as a way things aren't but could have been. This leads to the more standard kind of intension I started this discussion by considering. Given that the actual colourless, odourless stuff in oceans and lakes is H₂O, water refers to H₂O in all

¹⁰ Usually Chalmers talks about terms, rather than concepts, having primary/secondary intensions. We might think of the intension of a concept C as corresponding to the intension of a term expressing C. I talk of concepts rather than terms to fit in with the discussion of phenomenal concepts, but the claims of the paper could be easily expressed in terms of phenomenal terms rather than phenomenal concepts.

¹¹ Chalmers thinks of the worlds considered as actual as having 'centres' indicating the location of the speaker/thinker in the world. I shall ignore this subtlety to keep things simple.

possible worlds, and 'Water is H₂O' is true in all possible worlds. Therefore, we can say that the secondary intension of water delivers H₂O at all worlds, and the secondary intension of 'Water is H₂O' delivers the truth-value *true* at all worlds.

Evaluating the primary intension of concept or sentence is, by definition, an a priori matter. Locating water at world W considered as actual W does not require knowing which world is in fact actual.¹² We just search in W for the colourless, odourless stuff that falls from the skies and fills oceans and lakes. But evaluating the secondary intension very often requires empirical knowledge of relevant facts about the actual world. To know the secondary intension of water I need to know what the actual colourless, odourless stuff in oceans and lakes is.

In order to avoid trivialising the notion of a primary intension, we must restrict the kind of concepts which are deployed in the descriptions of the worlds used in the evaluation of a primary intension. To say that water has a primary intension is to say that, given enough information about a world considered as actual (and idealised rational faculties), we can work out what water refers to in that world. But this would be trivially true if the term 'water' was itself used in that description. We could then find the water by just looking for the stuff labelled 'water' in the description!

To avoid this worry, Chalmers stipulates that descriptions of worlds involve only indexicals and *semantically neutral* vocabulary. A semantically neutral term is one whose reference does not depend on which conceivable scenario turns out to be actual. Water is not semantically neutral, as its reference depends on how the actual world turns out, but zero and philosopher seem plausible candidates.

To turn to the argument, we can simplify things a little by thinking of premise 1 as the premise that zombies – entirely non-conscious physical duplicates of human beings – are conceivable (in fact, the conceivability of a zombie would be one instance of premise 1). Premise 2 relies on a general principle that if a sentence is conceivable, then it is '1-possible', which is to say its primary intension is true at some possible world: the sentence is true at some genuine possible worlds considered as actual (e.g. 'Water is XYZ', is conceivable, and hence it is true at some genuine possible world: the world where the stuff in oceans and lakes is XYZ). Premise 3 (given our simplification) tells us that a sentence asserting the existence of zombies has identical primary and secondary intensions, and so if it is true at some world considered as actual then it is true at that world considered as

¹² Chalmers defines the primary intension in terms of what could be known a priori given idealised rational faculties.

counterfactual. Premise 4 tells us that the truth of some zombie asserting sentence at some possible world considered as counterfactual is inconsistent with physicalism.¹³

Phenomenal transparency

I want to focus in this sub-section on premise 3 of the two-dimensional argument:

3. If $P \& \sim Q$ is 1-possible, then $P \& \sim Q$ is 2-possible

Chalmers grounds premise 3 in the thesis that the primary intension of a given phenomenal concept is identical to its secondary intension:

...the truth of premise 3 requires that both P and Q have primary intensions that coincide. In the case of Q , this claim is quite plausible... As Kripke noted, there does not seem to be the same strong dissociation between appearance and reality in the case of consciousness as in the cases of water and heat. Whilst it is not the case that anything that looks like water is water or that anything that feels like heat is heat, it is plausibly the case that anything that feels like consciousness is consciousness. So it is not clear that the notion of “pseudoconsciousness,” something that satisfies the primary intension of ‘consciousness’ without being conscious, is coherent. The same holds for other, more specific phenomenal properties. So there is a strong case that the primary and secondary intensions of phenomenal terms coincide.

Premise 3, then, is reliant on the following thesis:

Identity of Phenomenal Intentions: For any phenomenal concept, its primary intension is the same as its secondary intension.¹⁴

¹³ Chalmers uses the word ‘physicalism’ and ‘materialism’ interchangeably.

¹⁴ I am oversimplifying in two respects. As noted by Chalmers in the above quotation, premise 1 also requires that there is a no distinction between the primary and secondary intensions of P , which allows for the loophole of Russellian monism, which we will discuss below. However, what is crucial for Chalmers’ argument – modulo further qualifications I consider in footnote 17 – and for mine is that the primary and secondary intensions of phenomenal concepts are identical. Secondly, Chalmers also suggests an alternative way that the two-dimensional argument might go through, even if phenomenal concepts have different primary and secondary intensions. If $P \& \sim Q$ is conceivable, then its primary intension is true at some world W conceived as actual. Our world must differ from W , as the primary intension of $P \& \sim Q$ is false at our world. Given that W is physically indiscernible from our world, there must be some extra non-physical objects or property at either our world or W . P contains a ‘that’s all’ fact, which specifies that there is nothing more in W than the physical facts specified by P . Therefore, there must be some extra non-physical object or property present in our world, which distinguishes our world from W . On this style of argument, even if we don’t have a priori access to what is required for consciousness to be satisfied, we do have a priori access to what is required from the primary intensions of consciousness to be satisfied (for consciousness to actually refer at a world considered as actual). We could form a concept consciousness*, which denotes the property expressed by the primary intension of consciousness, and replace all discussion of ‘consciousness’ or ‘phenomenal properties’ in what follows with talk of ‘consciousness*’ and phenomenal* properties’, and then move from *Phenomenal* Transparency* to the

It is *Identity of Phenomenal Intentions* that gets us from zombies at worlds considered as actual, to zombies at worlds considered as counterfactual; the latter being required to refute physicalism (in footnote 17 I discuss Chalmers' suggestion for how to run the argument without relying on this premise). I want now to explore the implications of this thesis.

We can think of the secondary intension of as capturing what the concept requires of reality for its satisfaction. The secondary intension of water tells us that water requires of each metaphysically possible world the existence of H₂O: if H₂O exists at world W then water is satisfied at W, if H₂O does not exist at world W then water is not satisfied at W. Where the primary intension of a given concept differs from the secondary intension, we will not have a priori access to what that concept requires of reality for its satisfaction. In the case of water, we need to do empirical investigation in order to work out that water requires the existence of H₂O for its satisfaction. But if the primary intension of a given concept is identical to its secondary intension, given that the primary intension is a priori evaluable, the secondary intension will also be a priori evaluable. For such a concept, what it requires of the world for its satisfaction will be a priori accessible. Therefore, *Identity of Phenomenal Intentions* implies that it is a priori what phenomenal concepts require of the world for their satisfaction.

Call sentences which involve quantifiers rather than singular terms, and predicates which express phenomenal concepts, 'phenomenal sentences'. If it is a priori what phenomenal concepts require for their satisfaction, then plausibly it is a priori what phenomenal sentences require for their truth. If we can know a priori what is required from reality in order for the concept pain to be satisfied, then we can know a priori what is required for the sentence, 'There is something in pain' to be true. *Identity of Phenomenal Intentions*, therefore, gives us reason to believe the following thesis:

Phenomenal Transparency: It is a priori what phenomenal sentences require for their truth.

Each of the anti-physicalist arguments, at some stage, relies on a similar principle asserting our transparent understanding of the metaphysical demands of phenomenal concepts/sentences: Bealer's 'semantic stability', Nida-Rümelin's 'grasping', my 'transparency'. Indeed, in more recent work explicating the fundamental notions expressed by the two-dimensional framework, Chalmers talks explicitly in terms of phenomenal concepts being 'referentially transparent', in the sense that their extensions are knowable a priori.¹⁵

conclusion that cheap truth theory cannot account for truths about consciousness*. For the sake of keeping things simple, I shall ignore this complication in what follows.

¹⁵ Chalmers 2012: Excursus 14. Martine Nida-Rümelin (2007) gives a more detailed account of phenomenal transparency (defined on this page) in terms of the two-dimensional framework.

Whether explicit or implicit, it is Phenomenal Transparency which is at the heart of each of the anti-physicalist arguments. Without some such principle, it is hard to see how an epistemic gap between the physical and the phenomenal could have metaphysical significance.¹⁶ For if it is not a priori what phenomenal sentences require for their truth, then there will be an epistemic gap between the physical and the phenomenal *for precisely that reason*, even if those truth requirements entail physicalism. For example, suppose the sentence 'There is someone who feels pain' requires for its truth that there are c-fibres firing, but that that truth requirement is not a priori accessible. It follows that it is conceivable, but not possible, that there be c-fibres firing without pain. If we cannot access the truth requirements of phenomenal sentences a priori, then that fact in itself entails an epistemic gap between the physical and the phenomenal in a way that has no implications for whether or not physicalism is true.

There is good reason, then, to think that all the standard anti-physicalist arguments depend on a commitment to Phenomenal Transparency. Whilst I do not have space here to go into the details of each of the anti-physicalist arguments, I hope at least to have shown that premise 3 of Chalmers' argument depends on a commitment to Phenomenal Transparency.¹⁷

4. Anti-physicalist forms of cheap truth theory

It might be thought obvious that the anti-physicalist arguments threaten cheap truth theory. It is often supposed that the anti-physicalist arguments are trying to demonstrate that human mentality is fundamental, and the thesis that human mentality is fundamental seems inconsistent with a view which commits only to concrete simples (given certain other extremely plausible assumptions we will explore below). However, the anti-physicalist arguments are not trying to demonstrate that human mentality is fundamental; they are in general consistent with *non-physicalist forms of reductionism* about human mentality. This is made explicit in Chalmers' argument by its allowance for the loophole of *Russellian monism*.

¹⁶ One interesting exception is Joe Levine's (2014) recent defence of an *argument for the best explanation* against physicalism: Levine thinks the best explanation of the explanatory gap is that there is a metaphysical gap. The anti-physicalist arguments I am concerned with in this paper are attempts to demonstrate the falsity of physicalism a priori.

¹⁷ To be more precise, premise 3 of Chalmers' argument is dependent on the referential transparency of phenomenal concepts, which I have argued makes Phenomenal Transparency extremely plausible. Perhaps there is room for the cheap truth theorist to respond to the argument of this paper by accepting that phenomenal concepts are referentially transparent, but denying Phenomenal Transparency as I have defined it above (in terms of phenomenal sentences). This would be an interesting position to explore. The aim of this paper is to set up a debate which needs to be had, rather than explore every possible move in that debate.

Russellian monism, a view inspired by (independent) writings of Russell and Eddington from the 1920s,¹⁸ is currently enjoying a great deal of attention in the philosophical literature on the mind-body problem.¹⁹ There are two basic claims involved in the view. Firstly the claim that the physical sciences only characterise matter in terms of its structural or extrinsic nature, leaving us in the dark about the (more than structural) intrinsic nature which realises that structure. Let us call the properties in terms of which the physical sciences characterise matter 'pure physical' properties, and the (more than structural) intrinsic properties of matter which according to Russellian monism the physical sciences do not reveal to us 'deep' properties. Secondly, Russellian monists claim that it is the deep properties of matter which explain consciousness, and that it is the failure of the physical sciences to characterise this deep nature which accounts for the mind-body problem.²⁰

To put it in terms of Chalmers' two-dimensional framework, the Russellian monist holds that although the primary and secondary intensions of phenomenal concepts coincide, the primary and secondary intensions of physical concepts come apart. The primary intension of, say, mass picks out the pure physical nature of mass, whilst the secondary intension of mass picks out its deep nature. It is because the primary and secondary intensions of physical concepts come apart that zombies are conceivable. If we could conceive of matter in terms of its deep physical nature zombies would cease to be conceivable.

Is Russellian monism a form of physicalism? It depends on how we define physicalism. Some take physicalism to be the view that ideal physics exhaustively describes fundamental reality. Clearly Russellian monism is not a form of physicalism on such a definition. Others define physicalism as the view that physics is *referentially adequate*, in the sense that the fundamental facts, individuals or properties are those which are the subject matter of ideal physics, leaving it open whether or not ideal physics reveals their complete nature.²¹

In either case, the standard 'anti-physicalist' arguments do not rule it out, and for this reason the standard anti-physicalist arguments are not, in themselves, a threat to cheap truth theory. For we could construe Russellian monism as a form of cheap truth theory: the only items in the correct ontology are physical simples instantiating deep properties; states of affairs involving complex

¹⁸ Russell 1927, Eddington 1928.

¹⁹ See for example Pereboom 2011, and the essays collected in Alter & Nagasawa 2015.

²⁰ The general understanding of 'structure' in this context is roughly *what can be captured in vocabulary involving only mathematical, formal and nomic terms*. Russellian monists follow Russell (1927: 325) in arguing that there must be more to the nature of concrete reality than causal structure.

²¹ See Ney 2008 for a good survey of views on this aspect of the definition of physicalism.

arrangements of these simples make true sentences concerning human mentality. Most agree that this option is not ruled out by the anti-physicalist arguments.

Whilst not referring specifically to Russellian monism, Theodore Sider has also pointed out that anti-physicalist arguments are no threat to cheap truth theory. In his recent defence of micro-simplist cheap truth theory,²² he considers what he calls the 'Cartesian argument' against it:

1. I exist
2. I am not a simple entity
3. Therefore, there is at least one entity which is not simple.

Sider claims that although I may have Cartesian certainty that sentences concerning my existence are true, I have no good reason to think that such sentences need be made true by a single entity that has my conscious experience, a thesis he dubs 'metaphysical singularity':

Rejecting materialism about the world would not on its own support metaphysical singularity. Irreducible or supervenient mentality could consist of irreducible or nonsupervenient mental relations which relate many subatomic particles, rather than irreducible or nonsupervenient mental properties that are instantiated by single entities.²³

Whilst Sider is right that anti-physicalism is consistent with cheap truth theory, I will try to show in the next section that the key premises of the anti-physicalist arguments pose a threat to cheap truth theory of which Sider is unaware.

5: The phenomenal mirroring argument

Following J. R. G. Williams, we can think of cheap truth theories as advocating an anti-Quinian account of the truth requirements of sentences, or at least sentences in contexts.²⁴ If the sentence, 'There is a party at Bill's', only requires for its truth that there be people dancing, drinking, etc. at a certain location, then the truth of 'There is a party at Bill's' can consist in an indirect correspondence between the sentence and the people gathered at the relevant location. We can say that 'There is a party at Bill's' has 'indirect truth requirements', i.e. requirements that can be satisfied by entities other than those the sentence quantifies over, properties other than those expressed by the sentence's predicates, etc. Where a sentence (or a sentence in a context) has indirect truth requirements, the truth of that sentence may consist in an indirect correspondence with reality.

²² Sider 2012b. Actually, Sider is somewhat agnostic over whether we can get full blown truth on the cheap, or merely semantic correctness.

²³ Sider 2013: 29.

²⁴ Williams 2010.

However, there may be certain sentences, or certain sentences in certain contexts, which lack indirect truth requirements, and hence can only be satisfied by a direct correspondence with reality. Call such sentences 'mirroring sentences.' Sentences concerning God are plausible candidates for mirroring sentences. The sentence, 'God exists', seems to require for its truth nothing less than the existence of the entity it quantifies over.

In general, whether or not our sentences are mirroring sentences will not be a priori accessible. However – and here is where the premises of the anti-physicalist arguments start to become relevant – if we accept *Phenomenal Transparency*, then the truth requirements of phenomenal sentences will be a priori accessible. If the truth requirements of phenomenal sentences are a priori accessible, then it will presumably be a priori accessible whether or not those truth requirements are indirect, and hence whether or not phenomenal sentences are mirroring sentences.

I submit that, if it is indeed the case that the truth requirements of phenomenal sentences are a priori accessible, it is extremely plausible that phenomenal sentences are mirroring sentences. When we reflect on what would be required for 'There is something in pain' to be true, the only state of affairs that suggests itself as sufficient is *there being some entity that feels pain*, as a state of affairs which involves the entity quantified over by the sentence. Contrast with the sentence 'There is a party'. When we reflect on what would be required for this sentence to be true, it is apparent to us that it would be sufficient for its truth that there be people gathered together revelling, a state of affairs which does not involve the entity quantified over. In this way, it is apparent upon reflection that 'There is a party' has indirect truth requirements. However, in the case of phenomenal sentences, no such indirect requirements are apparent to reflection, even when we consider different contexts in which the sentence might be used.²⁵

Having said that, there is an option for those wanting to hold that there are a priori accessible indirect truth requirements of phenomenal sentences: give some kind of causal analysis of phenomenal sentences. An analytic functionalist may take it to be a priori that what is required for 'There is something in pain' to be true is for a certain causal role R to be realised, where this may involve a single state playing R, or may involve a number of entities (e.g. particles) acting in concert to play R. In so far as it is a priori that the truth requirements of this sentence may be satisfied by many things which aren't subjects of experience (e.g. non-conscious particles acting in concert to play a certain causal role), this would seem to constitute an a priori accessible indirect truth requirement of this sentence. An analytic behaviourist, or an analytic representationalist, may give a similar account. Theodore Sider attempts to give a functionalist account of the 'metaphysical truth

²⁵ Plausibly the content of a phenomenal concept does not vary with context.

conditions' (roughly what I am calling truth requirements) of some of our ordinary sentences, in the defence of cheap truth theory.²⁶

At this point the epistemic premise of the anti-physicalist arguments has relevance. For in zombie worlds things are arranged causally as the analytic functionalist believes to be required for the truth of 'There is something in pain', and yet in the zombie world 'There is pain' is not true. Therefore, if what is required for 'There is something to be in pain' to be true is a priori accessible, and the analytic functionalist is correct about what is required for 'There is something to be in pain' to be true, then zombie worlds would not be conceivable. Of course it follows that if zombie worlds are conceivable, and what is required for 'There is something in pain' to be true is a priori accessible, then the analytic functionalist must be wrong about what is required for 'There is something in pain' to be true.

Putting such causal analyses of phenomenal sentences on one side, I know of no other theory of phenomenal sentences which might yield indirect truth requirements of phenomenal sentences which are a priori accessible. Perhaps some might say that there are indirect truth requirements of phenomenal sentences, but we just haven't uncovered them yet. A few decades of searching for the truth requirements of knowledge claims has not yielded necessary and sufficient conditions for attributions of knowledge. Maybe it would be premature to infer from the fact that philosophers have not yet discovered a priori accessible indirect truth requirements of phenomenal sentences, to the fact that they don't exist.

However, the two cases are not relevantly similar. Even though philosophers have not managed to give a completely analysis of the concept of knowledge, in the sense of giving necessary and sufficient conditions for someone to know that p, we clearly have a rough idea of what is required for someone 'to know that p', and we are able to judge in particular cases not described using the verb 'to know' or related terms, whether or not those cases count as cases of knowledge. This makes us confident that the truth requirements of 'X knows that P' can be satisfied by state a state of affairs which can be specified without using the verb 'to know' or related terms.

But in the case of 'There is something in pain' – again assuming the falsity of causal analyses of that sentence – careful reflection does not yield even a rough idea of how entities other than the one quantified over by that sentence could make that sentence true; nor can we imagine a single state of affairs involving such entities which strikes us as sufficient for its truth. Careful reflection yields nothing that might sufficient for the truth of 'There is something in pain' other than the existence of

²⁶ Sider 2013. Actually, Sider is somewhat agnostic over whether we can get full blown truth on the cheap, or merely semantic correctness.

something that feels pain. Believing that there are a priori accessible indirect truth requirements for phenomenal sentences (other than those given by causal analyses), when no careful investigation yields no suggestion of their existence, is an unwarranted leap of faith.²⁷

One might be tempted to think that most sentences are such that there are no a priori accessible indirect truth conditions. But in general this need not be a problem for cheap truth theories. The cheap truth theorist could claim that 'There are tables' requires for its truth that particles be arranged table-wise, even though it is not a priori accessible that the sentence has these truth requirements; the truth requirements of sentences about tables may be determined by facts outside of what is a priori accessible. However, if we accept phenomenal transparency, then supposed indirect truth requirements of phenomenal sentences are either a priori accessible or they don't exist at all.

Let us put this all together. If we accept that zombies are conceivable, then we have very good reason to think that there are no a priori accessible indirect truth requirements of phenomenal sentences, as the conceivability of zombies rules out the only plausible candidates. If we further accept phenomenal transparency, the thesis that the truth requirements of phenomenal sentences are a priori accessible, then we have very good reason to think that the truth requirements of phenomenal sentences are not indirect, given that phenomenal sentences lack a priori accessible indirect truth requirements. The two crucial premises of the anti-physicalist arguments push us to the conclusion that phenomenal sentences lack indirect truth requirements.

We can put the argument as follows:

The phenomenal mirroring argument

1. If zombies are conceivable then phenomenal sentences lack a priori accessible truth requirements.
2. If phenomenal transparency is true, then: if phenomenal sentences lack a priori accessible indirect truth requirements, then phenomenal sentences lack indirect truth requirements.
3. Zombies are conceivable.
4. Phenomenal transparency is true.
5. Therefore, phenomenal sentences lack indirect truth requirements, i.e. phenomenal sentences are mirroring sentences.

²⁷ C.f. Theodore Sider's emphasis on the importance of a reductive theory offering 'toy' metaphysical truth conditions (Sider 2012: 116-8).

This would obviously be an unwelcome conclusion for cheap truth theorists. The commitment to *Moorean Truth* will not allow them to give up on the truth of phenomenal sentences; if anything *Moorean Truth* it is that there are the subjects of experiences we ordinarily take them to be, with more or less the conscious experience we pre-theoretically associate with them. If phenomenal sentences lack indirect truth requirements, then accepting their truth will require an ontological commitment to subjects of experience. But the cheap truth theorist adherence to *Minimal Commitments* entails that we can be ontologically committed to subjects of experience only if we take them to be simple, concrete entities.

Supposing the cheap truth theorist commits only to micro-level simples, she will be forced to identify the subjects of experience we pre-theoretically believe in with micro-level simples. This is clearly not an attractive view. Firstly, there is a concern about arbitrariness: it is hard to see what could determine which of the huge number of micro-level simples within my body the subject of my experience is identical with. Secondly, it is surely a Moorean truth that the subject of my experience causes much of my behaviour; if the subject of my experience is a single micro-level simple, it could not have the causal impact on my behaviour we pre-theoretically suppose that it has. Finally, it is plausibly a Moorean truth in its own right that the subject of my experience is not a micro-level simple. There are some anti-physicalists who adopt panpsychism, the view that micro-level simples have a crude form of consciousness,²⁸ but even on these views micro-level subjects do not instantiate the kind of experience we have a Moorean commitment to.

Suppose the cheap truth theorist commits only to the blobject. It is perhaps slightly more plausible to suppose that the universe instantiates my phenomenal properties than it is to suppose that a fundamental particle does. Following Horgan and Potrč, we might say that the universe instantiates my phenomenal properties R-wise, where R-wise is the spatio-temporal manner corresponding to my brain. One might have concerns about the intelligibility of phenomenal properties being instantiated in spatio-temporal manners, but let us put such concerns to one side.

The real trouble for blobjectivists comes when try to make sense of phenomenal sentences that quantify over multiple subjects of experience, e.g. ‘There is something that feels pleasure and no pain, and something that feels pain but no pleasure.’ Just as reflection does not reveal a state of affairs intuitively sufficient for the truth of ‘There is something that feels pain’, other than one that involves the existence of a pained subject, so reflection does not reveal a state of affairs intuitively sufficient for the truth of ‘There is something that feels pain but no pleasure, and something that feels pleasure but not pain’, other than the state of affairs of there being two subjects, one feeling

²⁸ Strawson 2006, Chalmers 2015, Goff 2015a.

pain but no pleasure and one feeling pleasure but no pain. It would not be correct to describe a state of affairs in which one thing feels both pleasure and pain – albeit in distinct spatio-temporal manners of instantiation – as a state of affairs in which that sentence is true.

It is not that there is anything incoherent with micro-level simples somehow constituting a subject of experience, or the cosmos in some sense grounding the existence of a subject. There are various anti-physicalist views according to which fundamental particles constitute subjects in virtue of their physical and non-physical properties. However, this kind of view is not cheap truth theory as I've defined it. Cheap truth theory is not just the view that simple concrete entities are metaphysically basic, but that they are the only entities that we are ontologically committed to. The only way to make this theory consistent with *Minimal Commitments* as I have defined it is to identify subjects of experience with immaterial souls. Although consistent with the letter of cheap truth theory, I take it that such a weighty ontological commitment is at odds with its spirit.

6: A debate that needs to be had

I hope to have made a case for the importance of linking up two areas of the philosophical literature which have so far existed in isolation: cheap truth theories in the metaphysics literature, and debates between physicalists and anti-physicalists about consciousness in the philosophy of mind literature. The anti-physicalist arguments in their standard formulations are not a threat to cheap truth theory, because – as Sider notes – the cheap truth theorist might adopt an anti-physicalist form of cheap truth theory. However, I have shown how the key premises of those arguments – Phenomenal Transparency and Zombie Conceivability – can be worked into an argument to the conclusion that phenomenal sentences are mirroring sentences, a thesis very hard to reconcile with cheap truth theory.

Of course those premises might be denied. By far the more popular response of contemporary physicalists to the anti-physicalist arguments involves denial of the referential transparency of phenomenal concepts. However, as physicalists have discovered, this is easier said than done. The opposite extremely to taking phenomenal concepts to be referentially transparent, is taking them to be referentially opaque: revealing *nothing* of the nature of their referents. In the 1990s and early 2000s many physicalist offered pure semantic externalist theories of phenomenal concepts which appeared to have that implication.²⁹ However, there is a growing consensus among both physicalist and anti-physicalists that this opposite extreme is deeply implausible. Conceiving of conscious states

²⁹ In the theories of Papineau (1993) and Perry (2001) phenomenal concepts are demonstratives; in Tye (1995, chapter 6) and Lycan (1996) they are indexicals; and in Loar (1990/97, 2003), Tye (2000, chapter 2), Carruthers 2004 and Perry 2001 they are recognitional concepts.

under phenomenal concepts seems to reveal something *substantial* or *significant* about their nature; physicalist Benedicté Veillet has dubbed the requirement that physicalists account for this in a way consistent with physicalism ‘the new challenge.’³⁰

A number of physicalists have recently tried to answer this challenge, offering accounts of phenomenal concepts such that they are *referentially translucent*: revealing something significant about conscious states, but without revealing their physical nature. For example, according to physicalist Robert Schroer, phenomenal concepts are descriptive concepts which reveal how conscious states are composed from more basic elements, where those more basic elements are denoted by referentially opaque sub-concepts; e.g. it is a priori that a particular shade of phenomenal orange is composed of phenomenal red hue and phenomenal yellow hue, where the sub-concepts phenomenal red hue and phenomenal orange hue are referentially opaque concepts referring to physical states.³¹ There are difficulties facing Schroer’s account, and others that have been proposed, but I will not explore these difficulties here.³²

In the context of cheap truth theory, we might understand the ‘new challenge’ as the need to give an account of phenomenal concepts such that they reveal something substantial about the nature of phenomenal properties, but without leading to the indirect truth requirements of phenomenal sentences being revealed a priori. Alternately cheap truth theorist may choose to embrace Phenomenal Transparency whilst revisiting the now unpopular position in philosophy of mind of denying the conceivability of zombies and giving a causal analysis of mental concepts.

Either way, the cheap truth theorist needs to give some account of phenomenal concepts which is both plausible and entails that the truth requirements of phenomenal sentences are indirect. It is not obvious how this can be done, as is indirectly evidenced by the difficulty contemporary physicalists have giving accounts of phenomenal concepts consistent with physicalism.

There is obviously lots more to be said, but the aim of this paper is to indicate that there is a debate here that needs to be had, rather than to examine every possible move within that debate. The Phenomenal Mirroring Argument poses a significant challenge to the cheap truth theorist; a response is due.

References

³⁰ Veillet forthcoming, see also Schroer 2010, Goff 2011, Diaz-Leon 2014, Balog 2012.

³¹ Schroer 2010.

³² See for example Goff 2015b.

Alter, T. & Nagasawa, N. (Eds.) 2015. *Consciousness and the Physical World: Essays on Russellian Monism*, Oxford University Press.

Balog, K. 2012. 'Acquaintance and the mind-body problem,' in C. Hill and S. Gozzano (Eds.) *New Perspectives on Type Identity: The Mental and the Physical*, Cambridge: Cambridge University Press.

Bealer, G. 1994. 'Mental properties,' *Journal of Philosophy* 91: 185-208.

Bealer, G. 2002. 'Modal epistemology and the rationalist renaissance,' in T. Gendler & J. Hawthorne (Eds.) *Conceivability and Possibility*, Oxford University Press.

Cameron, R. 2008. 'Truthmakers and ontological commitment,' *Philosophical Studies* 140: 1-18.

Cameron, R. 2010. 'Quantification, naturalness, and ontology,' in A. Hazlett 2010, *New Waves in Metaphysics*, New York: Palgrave Macmillan.

Chalmers, D. J. 1996. *The Conscious Mind: Towards a Fundamental Theory*, Oxford University Press.

Chalmers, D. J. 2002. *Philosophy of Mind: Classical and Contemporary Readings*, Oxford University Press.

Chalmers, D. J. 2009. 'The two-dimensional argument against materialism,' in B. McLaughlin (Ed.) *Oxford Handbook of the Philosophy of Mind*, Oxford University Press, 313-39.

Chalmers, D. J. 2012. *Constructing the World*, Oxford University Press.

Chalmers, D. J. 2015. 'Panpsychism and panprotopsyism,' T. Alter and Y. Nagasawa (Eds.) *Consciousness and the Physical World: Essays on Russellian Monism*, Oxford University Press.

Diaz-Elson, E. 2014. 'Do A Posteriori Physicalists Get Our Phenomenal Concepts Wrong?' *Ratio* 27: 1, 1-16.

Eddington, A. 1928. *The Nature of the Physical World*, Cambridge University Press.

Fine, K. 2001. 'The question of realism,' *Philosophical Imprint* 1.

Goff, P. 2011. 'A posteriori physicalist get our phenomenal concepts wrong,' *Australasian Journal of Philosophy* 89.

Goff, P. 2015a. 'Against constitutive Russellian monism,' in T. Alter and Y. Nagasawa (Eds.) *Consciousness and the Physical World: Essays on Russellian Monism*, Oxford University Press.

Goff, P. 2015b. 'Real acquaintance and physicalism,' in P. Coates & S. Coleman (Eds.) *Phenomenal Qualities: Sense, Perception and Consciousness*, Oxford University Press.

Horgan, T. & Potrč, M. 2008. *Austere Realism: Conceptual Semantics Meets Minimal Ontology*, MIT Press.

Sider, 2008/2012.

Jackson, F. 1982. 'Epiphenomenal qualia,' *Philosophical Quarterly* 32.

Jackson, F. 1986. 'What Mary didn't know,' *Journal of Philosophy* 83.

Levine, J. 1983. 'Materialism and qualia: The explanatory gap,' *Pacific Philosophical Quarterly* 64.

Levine, J. 2014. 'Modality, semantics and consciousness,' 167: 3, 775-784.

Nida-Rümelin, M. 2007. 'Grasping phenomenal properties,' in T. Alter & S. Walter (Eds.) *Phenomenal Concepts and Phenomenal Knowledge*, Oxford University Press, 207-36.

Pereboom, D. 2011. *Consciousness and the Prospects of Physicalism*, Oxford, New York: Oxford University Press.

Russell, B. 1927. *The Analysis of Matter*, Kegan Paul.

Schroer, R. 2010. 'What's the beef? Phenomenal concepts as both demonstrative and substantial,' *The Australasian Journal of Philosophy* 88, 505-22.

Strawson, G. 2006. 'Realistic monism: Why physicalism entails panpsychism,' in A. Freeman (Ed.) *Consciousness and its Place in Nature*, Imprint Academic.

Sider, T. 2009. 'Ontological realism,' in D. Chalmers, D. Manley & R. Wasserman (Eds.) *Metametaphysics* Oxford University Press.

Sider, T. 2012a. *Writing the Book of the World*, Oxford: Oxford University Press.

Sider, T. 2012b. 'Against parthood,' in K. Bennett & D. W. Zimmerman (Eds.) *Oxford Study in Metaphysics* 8, Oxford University Press.

Sider, T. 2013. 'Reply to Jonathan Schaffer,' in 'Symposia on *Writing the Book of the World*,' *Analysis* 73: 4, 751-770.

Veillet, B. Forthcoming. 'The cognitive significance of phenomenal knowledge,' *Philosophical Studies*.

Williams, J. R. G. 2010. 'Fundamental and derivative truths,' *Mind* 119: 103-41.

[Russell 1927; Eddington 1928; Feigl 1958; 1967; Maxwell 1979; Lockwood, M. 1989; Strawson 1994; 2003; 2006; Chalmers 1996; Griffen 1998; Stoljar 2001]